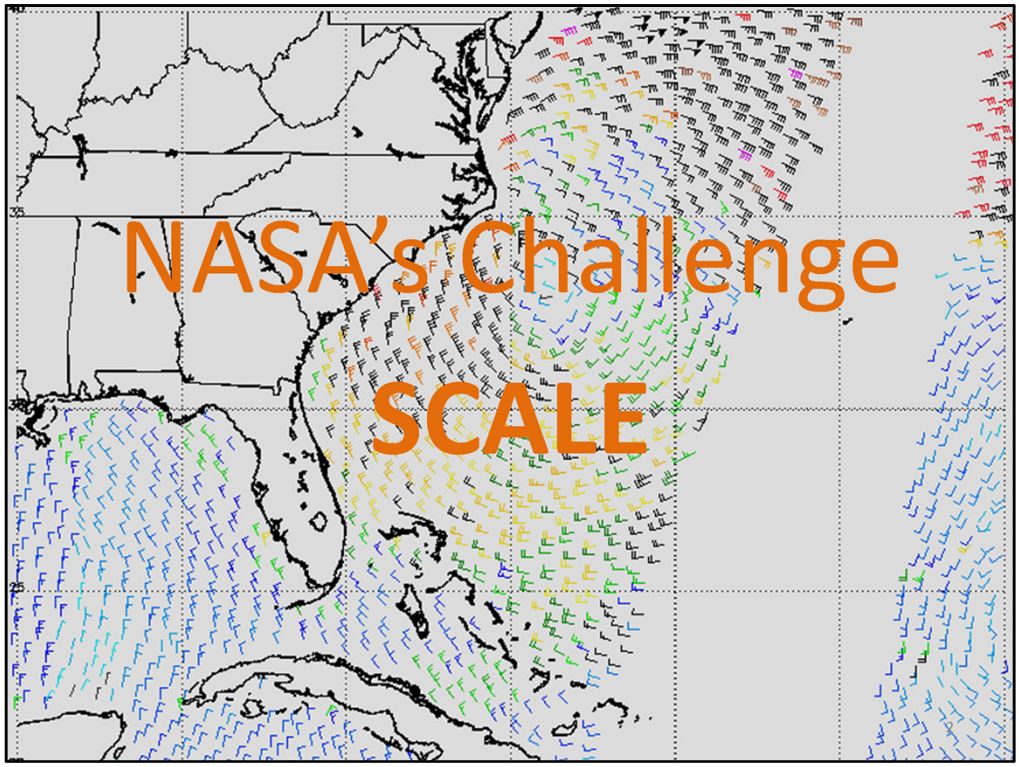
Themes:
Challenges of Big Data
- International
- Academia
- Federal Government & Industry
- NASA
- Earth Science Data



Although Unique, NASA still faces the similar challenges as other Big Data users... Federal, Industry and Academia

NASA's "Big Data" challenges go beyond the stereotypical "Big Data" problem-space -- involving warehousing and mining of unstructured transactional business data. This should be no surprise to the ESIP community. Many of you have significant day-to-day experience with Big Data and may be ahead of NASA on the innovation curve.

Not to say that we don't have unstructured data, but our key assets are our Science Data Sets – especially our earth science. Many of our "big data" sets are described by significant metadata, but on a scale that challenges current and future data management practice.

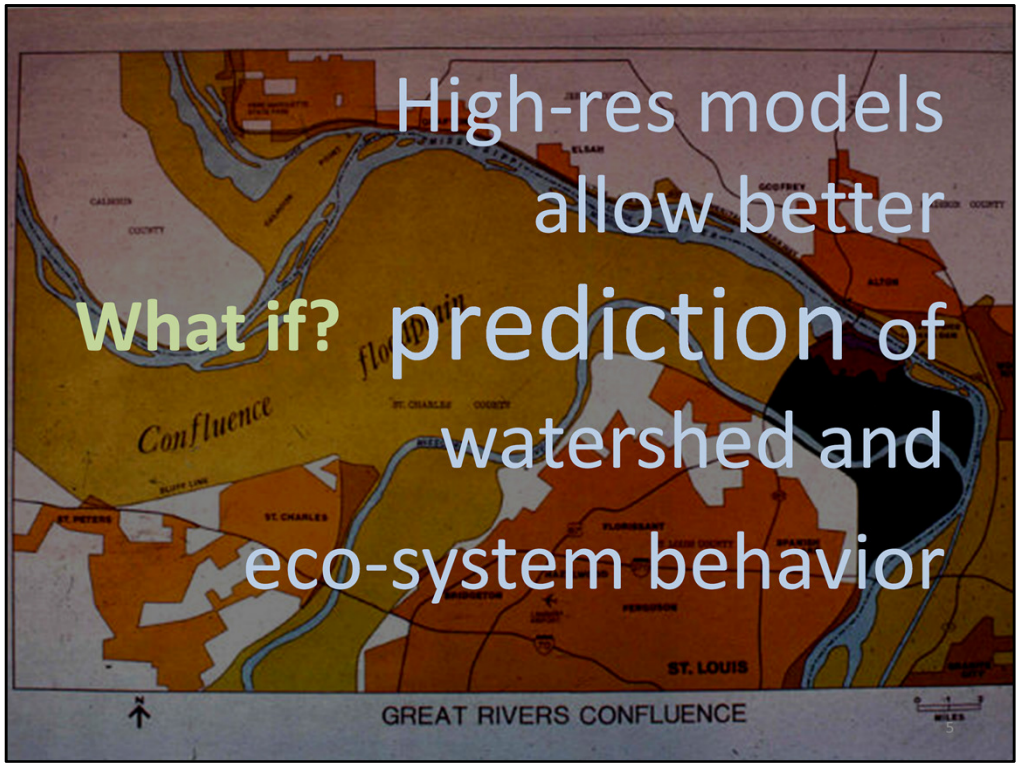
TRMM Satellite Data

Getting the right data to the right people, for decision support in real time.

NASA's data requirements demands collaboration beyond our current working relationships – to include leveraging capabilities and resources across the Federal Government, Industry, Academia, and NASA Centers to push the envelope toward new capabilities.

Making real-time data available to disaster workers. Where are the hardest-hit areas, how can we get in to help? What will we need to bring with us? How is the situation changing?

Sendai Japan, March 14, 2011 download GeoTIFF file (12 MB, TIFF, 3600x2400)

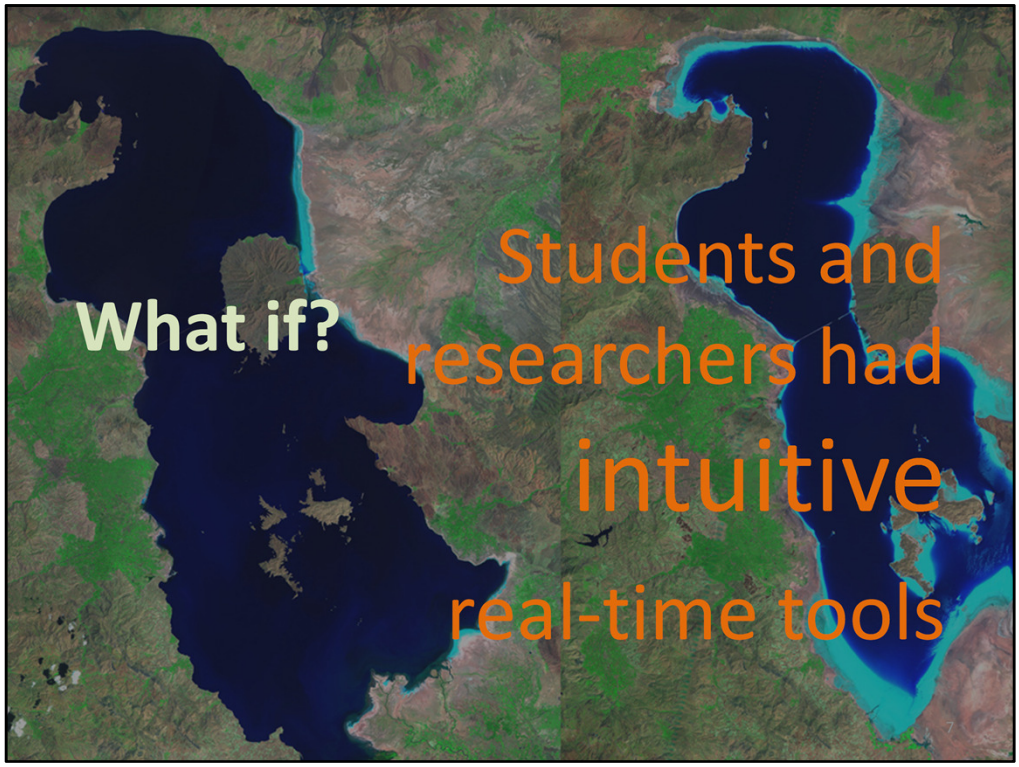


High stress eco systems respond to so many variables… Accuracy matters.
What if the data sets could help to predict their behavior?
Through detailed models we can help build ecological resilience and sustainability.

Map of Missouri and Mississippi river confluence around St Louis, Missouri

Should I buy this house or that? What kind of risk am I taking by wanting to build here? Using the latest data, help to give the answers to “Why?” Incredibly valuable to urban planners and civil engineers.
Insurance companies can use up-to-date and relevant data, reflecting true risk.

Spring 2011 Floods (Springfield) EO-1



With more intuitive real-time tools, researchers can easily 'see' the phenomena and students can quickly grasp/master the knowledge.
Help the educational community to tackle real problems, on a scale that they can handle.
Training the next generation of researchers through interdisciplinary tools and data with a lower burden.


Cloud services can help.

The dialogue must begin and end with Data Sensitivity

Challenges of Moving to "the Cloud"

- Initial infrastructure resource **$$$** investment is potentially significant
  - Equipment acquisition
  - Software retooling to run efficiently in the cloud
  - IT Security Costs
    - Public/Private
- Initial "risks" are high – **not everything will work**
  - Some cloud technologies may not develop into production-ready capabilities
  - Some science data experiments may not be well suited to cloud technologies
- New **Skill Sets** Required
  - Creating applications in the Cloud
  - migrating legacy applications to the Cloud
  - Programming skills, performance management techniques still need to be learned
  - Agility in the cloud will be an important skill

The rewards can be quite fruitful. The Proof of concept efforts…
Help to define how deep the rewards can be.
What the cost advantages can be.
Understand what duplication can be mitigated.

Initial resource investment is significant, however will save money in the long term, and save HPC resources for highly complex models

Optimizing use of cloud resources is a “new” skill – a skill that once gathered will be available for other institutional projects

You get more than just the capability at the end.

Leveraging Collaboration
Within the organization private cloud, outside the organization public cloud.
Share experience and findings, lessons learned, above all, best practices

Complementing traditional computing
Cloud technology fosters collaboration between data providers and compute providers.
Cloud is adaptable complement to HPC
Also, extends the capability to the average scientists at universities, eventually even citizen scientists.

Discovery Cluster

Form collaboration opportunities across organizational boundaries:

Shared risk/shared technical skills to build shared knowledge and experience… and relationships

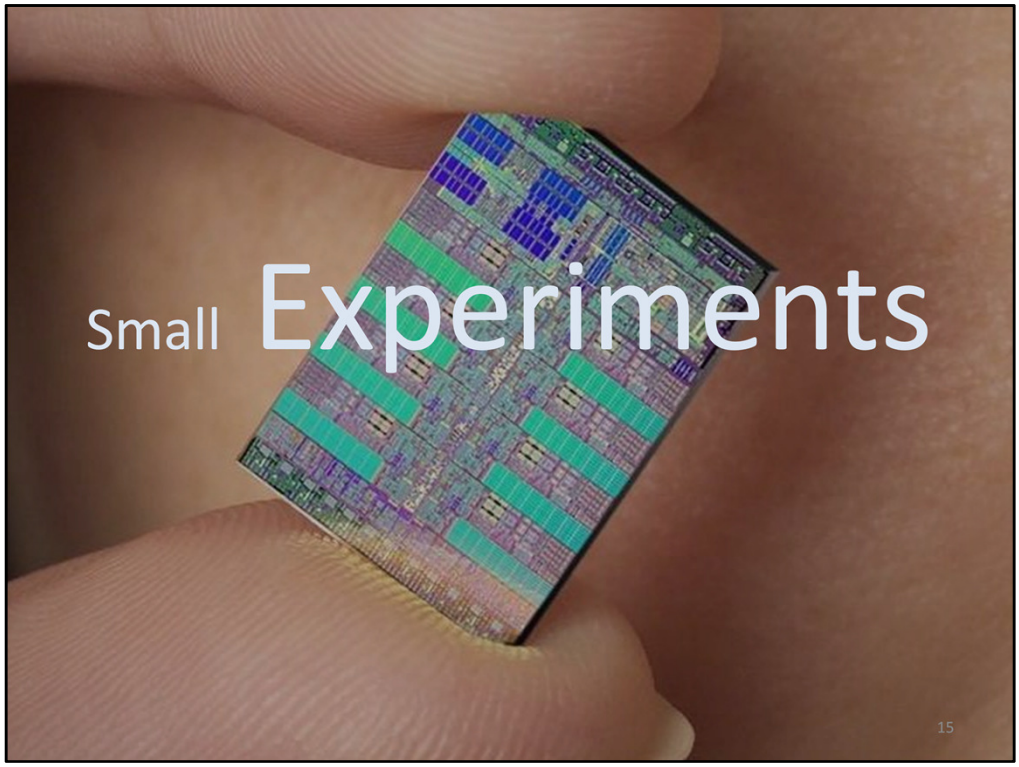
Reduces the entry barriers by limiting startup cost and sharing performance risks

Reach within your organization as well as outside

We want to get to a sustainable operational state.

New designs for data architectures, including clouds….

Collaborative Sustainable outcomes are the ideal outcomes.
Policy drives the approach through Proof of Concept efforts



Start with “small experiments” – well defined and measured

Builds experience and relationships

Examples of this can be found in NASA Goddard’s internal experiments

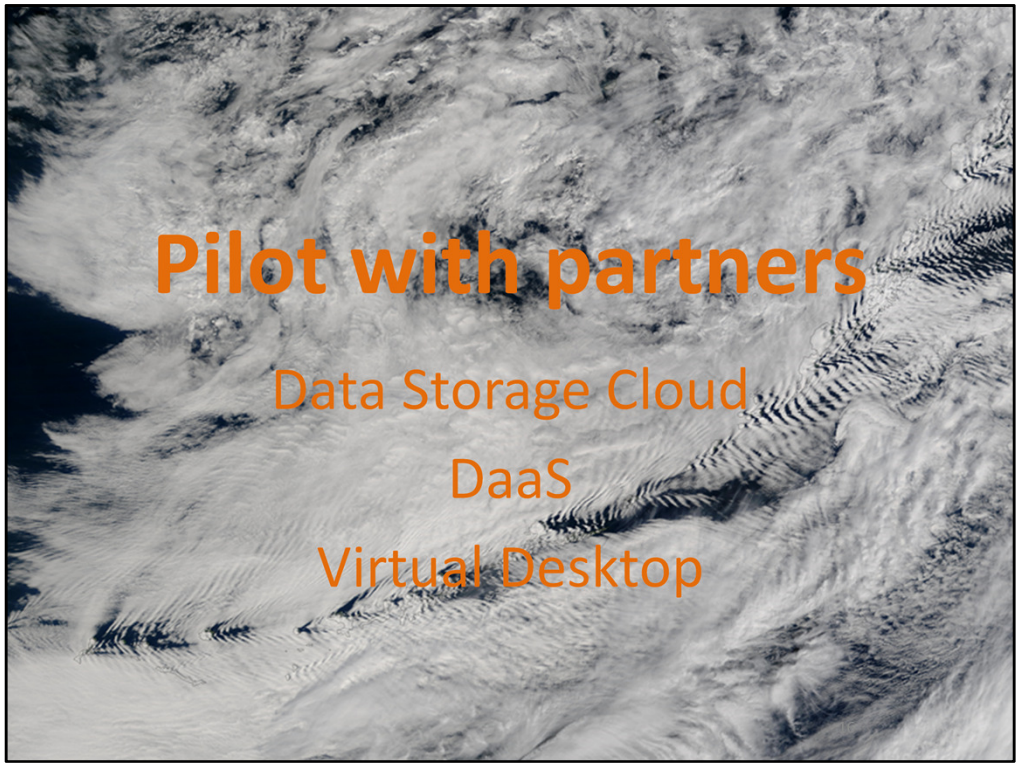
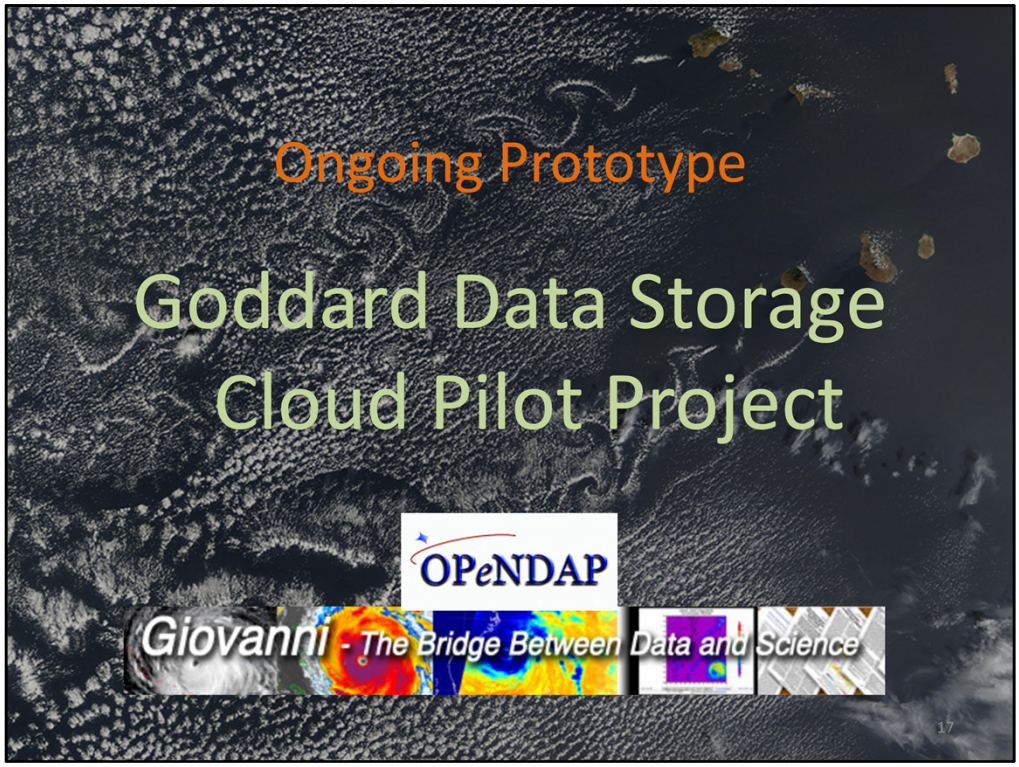


Image of the Aelutian islands by TERRA MODIS



Goddard Data Storage Cloud Pilot Project

To further explore and address the Big Data problems, a Proof of Concept is formed among different Goddard units. OpenDAP and Giovanni will be ported into a Goddard Data Storage Cloud to test

- End-to-end cloud solution vs. traditional client-server solutions
- Storage (volume & speed) bottlenecks
- Communication(networking) bottlenecks
- Processing (e.g., spatiotemporal subsetting and format/projection transforming) bottlenecks
- On demand support to Big Data access

Image from TERRA MODIS

How can we begin to think of our data sets as a greater Data as a Service
Prepare for a NASA Data as a Service (DaaS)

- Establish sustainable collaborations among difference units/organizations to share resources (people, environment, tools)
- Form a requirements, R&D, solutions, feedback cyclic process for addressing Big Data challenges
- Facilitate releasing the potential of Big Data
- Provide an elastic and pay-as-you-go solution for Big Data

Many needs, many data sets

Working Cloud Beyond Earth Science Data…

Goddard Persistent Virtual Desktop

- Working with IBM to pilot a persistent virtual desktop
- Goddard Heliophysics community working with ITCD to explore possibilities for science virtual desktop
- Could lead to the development of suites of collaborative science analysis applications as a service in the cloud, much the same way that Google Docs provides a collaborative office suite in the cloud

Image: SDO

Across Federal Government

National Science Foundation soliciting proposals for “Core Techniques and Technologies for Advancing Big Data Science & Engineering”

Industry

Using Space Act Agreement to work with IBM on Persistent Virtual Desktop

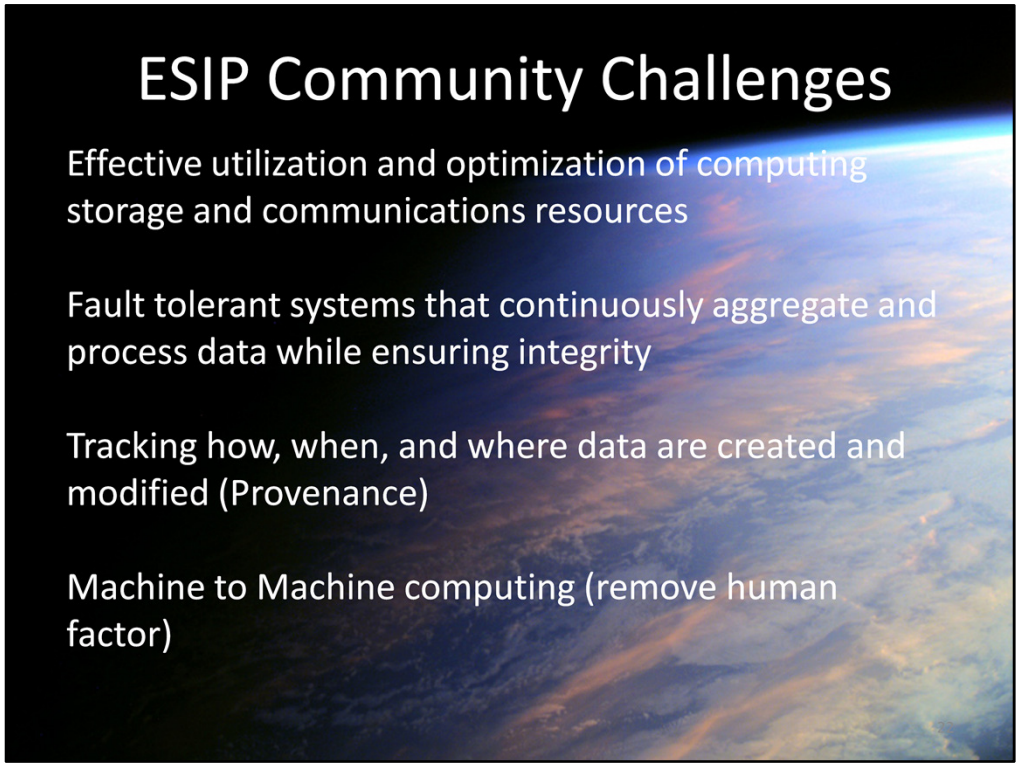
Microsoft has offered to beta test new virtual machine environments to support larger data volumes

ISS Photo

Work to leverage outside resources

- NASA resources are tight and risk tolerance is low
- Working to leverage outside resources, such as industry and academia to pilot innovation efforts to solve NASA's big data challenges
  - Within NASA
    - Potential for additional DSC experiments as team builds successes
    - Other teams at GSFC are interested in leveraging cloud technology
    - Other NASA Centers looking at data storage capability
  - Outside NASA… You?

Grand Canyon, Terra Satellite (composite 3-D view)



I know that there is a good bit of knowledge in these areas. The community can benefit from the sharing of experience in these areas to facilitate sustainable leveraging of Big Data
Leverage NASA's data in a way we never have in the past

Challenge to industry and academia

More opportunities exist between Federal Partners, NASA, industry, and academia… need to expand those relationships

If there are things that you can contribute that will make our offering better… Our work solving these problems should also be collaborative

ISS Photo


National Aeronautics and Space Administration
Questions?
NASA
If you want more information, please reach out:
Myra Bambacus myra.j.bambacus@nasa.gov
Lon Gowen lon.d.gowen@nasa.gov
Goddard
it